QUANTUM: Jurnal Inovasi Pendidikan Sains, Vol. 15, No. 2, 197-209
p-ISSN : 2086-7328 e-ISSN : 2550-0716
Website: https://ppjp.ulm.ac.id/journal/index.php/quantum/index
DOI: http://dx.doi.org/10.20527/quantum.v15i2

ANALYSIS OF THE VALIDITY AND RELIABILITY OF A CRITICAL THINKING SKILLS INSTRUMENT ON THE TOPIC OF WAVE-PARTICLE DUALITY USING RASCH MODEL

Tarpin Juandi[1,2], Ida Kaniawati[1*], Achmad Samsudin[1], Lala Septem Riza[1], Susilawati[3], Sapiruddin[2]
[1]Universitas Pendidikan Indonesia
Jl. Dr. Setiabudhi No. 229 Bandung 40154. Jawa Barat, Indonesia
[2]Universitas Hamzanwadi
Jl. TGKH. M. Zainuddin Abdul Madjid 132 Pancor Lombok Timur, Indonesia
[3]Universitas Mataram
Jl. Majapahit No. 62 Mataram 83115. Nusa Tenggara Barat, Indonesia
* corresponding author: kaniawati@upi.edu

Abstract. This study aims to analyse the validity and reliability of a critical thinking skills instrument on the topic of wave-particle dualism using the Rasch model. Data collection was carried out by administering a critical thinking skills test to students enrolled in a modern physics course. A total of 36 students from a university in West Nusa Tenggara participated. Data analysis was performed using the Rasch model through the Winsteps 4.6.1 software. The results indicated that the instrument is valid and reliable. The instrument's validity was tested by examining the data's fit to the Rasch model through infit and outfit MNSQ values and ZSTD, all of which were within the expected acceptance range. Construct validity was analysed through standardized residual variance, showing that the Rasch model can explain most of the variance in the data. Similarly, the instrument's reliability showed that the item reliability was in the very good category (0.93) and the person reliability was in the moderate category (0.68), with a Cronbach's alpha value of 0.86, indicating very good internal reliability. These findings confirm that the Rasch model is effective in assessing and improving the quality of critical thinking skills evaluation instruments in the context of modern physics education.

Keywords: critical thinking, modern physics, rasch model, reliability, validity

INTRODUCTION

Critical thinking plays an important role in modern physics for various reasons. Modern physics deals with highly abstract concepts, such as quantum mechanics and general relativity, which require deep understanding and proper application (Arabatzis, 2017; Chiang et al., 2015; Gubsky et al., 2019). Additionally, these theories often involve complex mathematics, so critical thinking is needed to ensure calculations are performed correctly and their theoretical implications are fully understood. In the context of experiments and observations, critical thinking is necessary to design valid and repeatable experiments as well as to carefully analyse data (Alves and Santos, 2021; Bao and Koenig, 2019). This includes separating signals from noise, recognizing patterns, and identifying anomalies that might indicate errors or new discoveries. Modern physics often intersects with other disciplines such as chemistry, biology, and materials science, so critical thinking is needed to integrate concepts from various fields (Hermann, 2020; Sun and Latora, 2020). Moreover, some

Published by: Program Studi Pendidikan Kimia FKIP Universitas Lambung Mangkurat
How to Cite: Juandi, T., Kaniawati, I., Samsudin, A., Riza, L.S., Susilawati & Sapirudin. (2024). Analysis Of The Validity And Reliability Of A Critical Thinking Skills Instrument On The Topic Of Wave-Particle Duality Using Rasch Model. *QUANTUM: Jurnal Inovasi Pendidikan Sains*, 15(2), 197–209. http://dx.doi.org/10.20527/quantum.v15i2.19996.

Article History:
Received : 01-04-2024, *Accepted* : 22-08-2024, *Published* : 24-10-2024

discoveries in physics have significant ethical and social implications, and critical thinking helps scientists consider the impact of their research and make responsible decisions.

In the context of the wave-particle duality concept, critical thinking skills are highly required. Wave-particle duality is a concept that shows that every particle or quantum can be exhibited as either a wave or a particle, depending on how it is observed (Sinaga and Feranie, 2017; Verawati et al., 2022). To understand and apply this concept, critical thinking plays an important role in several aspects. First, wave-particle duality requires a deep understanding of two seemingly contradictory models: wave properties and particle properties. Physicists must use critical thinking to identify situations where an entity is better explained as a wave or as a particle (Tarchi and Mason, 2019; Trilling and Fadel, 2009). For example, electrons exhibit particle properties when they are detected as points on a screen in a cathode ray tube experiment, but they exhibit wave properties when they undergo interference and diffraction in a double-slit experiment.

Secondly, understanding wave-particle duality also requires the ability to evaluate and interpret complex quantitative experiments. Experiments such as the photoelectric effect explained by Einstein or the double-slit experiment conducted by Thomas Young require critical thinking to interpret their results and realize the profound implications for fundamental concepts in physics (Alves and Santos, 2021; Yu et al., 2022). Furthermore, a high level of critical thinking is necessary to develop and evaluate theories that support this duality concept. Quantum theory, which underpins wave-particle duality, requires a critical approach to examine its theoretical accuracy and consistency, as well as its alignment with experimental evidence (Connor et al., 2018; Daineko et al., 2016). This includes understanding the Schrödinger equation, Heisenberg's uncertainty principle, and other concepts that underpin quantum mechanics.

Therefore, modern physics requires evaluation instruments that can measure not only lower order thinking knowledge but also students' critical thinking abilities. High-order thinking instruments are important to ensure that the learning process can facilitate the development of high-order thinking skills, which are essential for understanding complex physics concepts (Arabatzis, 2017; Kanim, 2020; Roza and Festiyed, 2023). In this context, valid and reliable evaluation becomes essential to identify and improve the quality of students' learning and understanding of the taught material. Technological advancements in psychometrics, such as the Rasch model, offer a more robust approach to measuring the validity and reliability of educational evaluation instruments. The Rasch model, which is part of item response theory (IRT), allows for a more in-depth analysis of item performance and respondent ability (Ling Lee et al., 2021; Sumintono and Widhiarso, 2015). It provides a more objective and quantitative framework for assessing the quality of instruments and ensuring that they truly measure what they are intended to measure.

The main problem in this research is the lack of valid and reliable evaluation instruments to measure critical thinking skills in modern physics courses. This is especially true for topics such as black body radiation, the X-ray Compton effect, the photoelectric effect, and the Bohr model for the hydrogen atom, or more generally, the topic of wave-particle duality (Alves and Santos, 2021; Sun and Latora, 2020). Many existing instruments are often not thoroughly tested for their validity and reliability, resulting in potentially inaccurate and unreliable evaluation outcomes.

Additionally, limitations in traditional analysis methods often fail to reveal in-depth information about individual item performance and respondent ability (Wei et al., 2020; Widhiarso and Sumintono, 2016). A common solution applied in various studies is the use of the Rasch model to analysed response data. The Rasch model allows researchers to conduct a more comprehensive and detailed analysis, including the ability to identify misfitting or biased items and to assess the internal consistency of the instrument (Planinic et al., 2019; Samsudin et al., 2023). Thus, the use of the Rasch model can enhance the accuracy and reliability of evaluation results, as well as provide better insights into how the instrument functions in different contexts.

The Rasch model has been widely used in various studies to evaluate and improve measurement instruments in the field of education. Research by Bond and Fox (2015) shows that the Rasch model can be effectively used to assess the validity and reliability of instruments in a more objective manner compared to classical methods. In this study, the Rasch model allows for the identification of poorly functioning items and provides useful information for instrument revision. Another study by Boone and Staver (2020) applied the Rasch model in the context of evaluating critical thinking skills. Their study results indicate that the Rasch model can identify biased items and improve the internal consistency of the instrument, ensuring that the instrument can be used more effectively to measure critical thinking skills. They also emphasize the importance of individual item analysis to understand how each item contributes to the overall construct of the instrument. In the context of physics education, a study by Engelhard (2013) applied the Rasch model to evaluate instruments used in modern physics courses. The findings of this research show that the Rasch model not only enhances the validity and reliability of the instruments but also provides deeper insights into students' understanding of complex physics concepts. This helps in identifying areas needing improvement in the curriculum and teaching methods.

Although the Rasch model has been widely used in various educational contexts, there is still a gap in research applying this model to measure critical thinking skills, particularly in modern physics courses. Many previous studies have focused more on evaluating basic cognitive knowledge without considering the dimension of higher-order thinking skills (Maryani et al., 2021; Nitriani et al., 2022). Additionally, many instruments used in these studies have not undergone rigorous validation processes using the Rasch model, so the results obtained may be less accurate and reliable. Furthermore, most research has focused on one aspect of validity or reliability alone, without integrating both comprehensively (Hadzhikolev et al., 2020; Rahmayanti et al., 2022). This highlights the need for more in-depth and holistic research that not only evaluates but also improves the measurement instruments for critical thinking skills in the context of modern physics courses.

This study aims to evaluate a valid and reliable instrument to measure critical thinking skills in modern physics courses using the Rasch model. The novelty of this research lies in its comprehensive application of the Rasch model to not only assess but also enhance the quality of evaluation instruments in the context of physics education. The scope of this study includes data collection from students enrolled in modern physics courses and data analysis using the Rasch model. The results of this research can make a significant contribution to the field of physics education by providing a more accurate and reliable evaluation tool to measure students' critical thinking skills.

RESEARCH METHOD

This research employed a descriptive quantitative method with Rasch analysis to measure the validity and reliability of critical thinking instruments in modern physics lectures. Data collection techniques in this study were carried out by administering a critical thinking skills test to 36 students participating in modern physics lectures at a university in West Nusa Tenggara. Students were randomly selected to ensure data representativeness. Before data collection, the instrument was developed based on relevant literature (Anselmi et al., 2019; Juandi et al., 2024; Mulhayatiah et al., 2022). The critical thinking skills instrument was developed based on critical thinking skill indicators by Nitko and Brookhart (2014). This instrument consists of several parts that cover various aspects of critical thinking, such as analysis, evaluation, inference, and reflection (Laliyo et al., 2022). Data analysis techniques use the Rasch model with Winsteps 4.6.1. Rasch analysis was chosen for its ability to provide a more in-depth assessment of item performance and respondent ability, as well as to identify and correct items that are inappropriate or biased (Bond and Fox, 2015; Boone and Staver, 2020; Ling Lee et al., 2021).

The parameters measured in this study include item difficulty level, respondent ability, and the internal consistency of the instrument. Item difficulty level is measured based on how many students answered the item correctly, while respondent ability is measured based on their performance on the overall instrument. The internal consistency of the instrument is measured using the reliability coefficient generated from the Rasch model analysis. This analysis includes examining the infit and outfit mean square to identify misfitting items, analyzing the item characteristic curve (ICC) to evaluate item performance, and calculating the reliability coefficient to assess the instrument's internal consistency (Handayani et al., 2023; Susbiyanto et al., 2024; Widhiarso and Sumintono, 2016).

Several Rasch model terms that need attention include Infit (Inlier-Sensitive Fit), which is sensitivity to misfit affecting person ability or item difficulty equally. Outfit (outlier-sensitive fit) is sensitivity to misfit influenced by outliers or extremely high values. MNSQ (mean square) indicates the mean square of deviations, ZSTD (standardized Z-score) represents the standardized deviation value of the deviations, and the point measure correlation value refers to the correlation between the item and the overall measure. The criteria for each parameter are: outfit MNSQ value between 0.5 to 1.5, outfit ZSTD value between -2.0 to +2.0, and point measure correlation value between 0.4 to 0.85 (Andrich and Marais, 2019; Boone and Staver, 2020; Planinic et al., 2019). After the analysis is conducted, the results are used to revise the instrument and ensure that it is valid and reliable for measuring critical thinking skills in modern physics lectures.

RESEARCH RESULTS AND DISCUSSION

An analysis of the critical thinking instrument trial data has been conducted, and the results presented in this section relate to the reliability and validity of the instrument. Regarding reliability, this analysis outlines person reliability, item reliability, and Cronbach's alpha values. As for validity, both instrument-level validity and item-level validity are described. Instrument-level validity includes data fit to the model (whether the data fits the model) and construct validity. Item-level validity encompasses how much, how precise, how good, and differential item functioning.

Reliability

In Rasch model analysis, reliability values are not represented by a single accumulated value but are presented as three separate values: person reliability, item reliability, and Cronbach's alpha. This provides a more complex interpretation of an instrument. The summary statistics from the Rasch model analysis can be seen in Table 1.

Table 1. Summary statistics by Rasch modelling

	Mean	SD	Separation	Reliability	Alpha Cronbach
Person	-0.24	0.71	1.47	0.68	0.86
Item	0.00	1.06	3.54	0.93	

Table 1 shows an average person value of -0.24, which indicates the average respondent ability measured in logits. The negative value signifies that the average respondent ability is below the average item difficulty (Boone and Staver, 2020). The standard deviation for each person is 0.71, which indicates the spread of respondent abilities. The higher the SD value, the greater the variation in respondent abilities (Planinic et al., 2019). The separation value is 1.71, which indicates how well respondent abilities can be separated into different groups. A value greater than 1 signifies a significant difference between higher and lower respondent abilities. The person reliability value is 0.68, which indicates the consistency of measuring respondent abilities. This value falls within the moderate category. Generally, a value above 0.70 is considered good, but a value of 0.68 is still acceptable (Widhiarso and Sumintono, 2016).

Table 1 also shows the measurement results for the test items. The average item value is 0.00, which indicates the average item difficulty measured in logits. Since the average item difficulty is 0.00, it means that the item values have been calibrated to centre around zero. The standard deviation for items is 1.06, indicating the spread of item difficulties. The higher the SD value, the greater the variation in item difficulty. The separation value is 3.54, which indicates how well the item difficulties can be separated (Anselmi et al., 2019; Lin et al., 2018). A value greater than 2 signifies that the items have significant differences in difficulty, and a value of 3.54 indicates that this differentiation is very good. The item reliability value is 0.93, which shows the consistency of measuring item difficulties. A value of 0.93 indicates excellent reliability, meaning the measurement of item difficulties is very consistent. On the other hand, the Cronbach's alpha value is 0.86, which is a measure of the internal reliability of the overall test. A value of 0.86 indicates very good reliability. Generally, values above 0.70 are considered good, and values above 0.80 are considered very good (Andrich and Marais, 2019; Sumintono and Widhiarso, 2015). Overall, Table 1 shows that the average respondent ability is slightly below the average item difficulty, there is considerable variation in respondent abilities and item difficulties, the test has good capability in separating respondent abilities and item difficulties, measurement reliability is good for items and moderate for respondents, and the overall internal reliability of the test is very good.

Instrument-Level Validity

The validity at the instrument level was verified in two stages. The first is confirming the fit of the data to the model (whether the data fit the model). The second is examining the construct validity of the instrument. The data fit with the Rasch model was assessed based on the criteria of the infit MNSQ and outfit MNSQ values in the summary statistic table. The analysis results for this parameter are shown in Table 2.

Table 2. Person and item fit with the Rasch model

	Infit		Outfit	
	MNSQ	**ZSTD**	**MNSQ**	**ZSTD**
Person	0.97	-0.08	0.96	-0.11
Item	1.02	-0.24	1.00	-0.26

Based on Table 2, the person Infit MNSQ and outfit MNSQ values are 0.97 and 0.96, respectively, which are very close to 1.00, indicating that the response data from students fit the Rasch model. Similarly, the person infit ZSTD and outfit ZSTD values of -0.08 and -0.11, respectively, show very small standard deviations, meaning there is no significant misfit of the data to the model. For items, the infit MNSQ value of 1.02 is very close to 1.00, and the outfit MNSQ value of exactly 1.00 indicates that the item data fits and is very consistent with the Rasch model. The item infit ZSTD and outfit ZSTD values of -0.24 and -0.26, respectively, show small standard deviations, meaning there is no significant misfit of the data to the model. Therefore, the data show that both students (persons) and items have a good fit with the Rasch model, from both the infit and outfit perspectives. The MNSQ values close to 1.00 and ZSTD values close to 0 indicate that the data fit the Rasch model very well.

Construct validity is assessed based on the values of raw variance explained by measures, unexplained variance in the 1st and 2nd contrasts, and eigenvalues. These variable values can be seen in Table 3.

Table 3. Standardized residual variance in eigenvalue units and item information units

Variabel	Eigenvalue	Observed	Expected
Raw Variance Explained by Measures	18.4281	52.0%	52.5%
Raw Variance Explained by Persons	2.8018	7.9%	8.0%
Raw Variance Explained by Items	15.6263	44.1%	44.5%
Unexplained Variance in 1st Contrasts	3.5624	10.1%	21.0%
Unexplained Variance in 2nd Contrasts	2.4466	6.9%	14.4%

The raw variance explained by measures (a combination of person and item) is 18.4281 or 52.0%, which is very close to the expected value of 52.5%. This indicates that the Rasch model is quite effective in explaining the variance in the data (Andrich and Marais, 2019). Next, the raw variance explained by persons (individual differences) is 2.8018 or 7.9%, approaching the expected value of 8.0%. This suggests that individual differences significantly contribute to the data variance. Finally, the raw variance explained by items is 15.6263 or 44.1%, which is very close to the expected value of 44.5%. This indicates that the differences in item difficulty significantly contribute to the data variance (Andrich and Marais, 2019; Boone and Staver, 2020).

In the Rasch model analysis, there are unexplained factors that contribute to the data variance, which ideally should not exceed 15%. Based on Table 3, the unexplained variance in the 1st contrast has a value of 3.5624, or 10.1%, indicating the presence of a significant additional factor not explained by the Rasch model. Furthermore, the unexplained variance in the 2nd contrast has a value of 2.4466 or 6.9%, suggesting the existence of a second factor not explained by the model. These values indicate variance not accounted for by the instrument, contributing to the overall data variance.

These results indicate that the Rasch model is quite effective in explaining the variance in the data, with more than 52% of the variance explained by measures (persons and items) (Andrich and Marais, 2019; Lin et al., 2019). The Rasch model explains the variance in the data quite well, indicating the percentage of total variance in the data that can be explained by the Rasch model. However, there is some variance not explained by the model (a total of 48%), with some contrasts indicating the presence of additional factors influencing the data. This suggests that although the Rasch model fits well, there is still room for improving the model or considering additional factors in the analysis (Andrich and Marais, 2019; Boone and Staver, 2020; Widhiarso and Sumintono, 2016).

Item-Level Validity

Item-level validity is assessed based on how much, how precise, how good, and differential item functioning. How much refers to the item measure value, indicating the level of item difficulty, and how precise refers to the S.E. value, indicating the precision of the item. To determine how much and how precise, refer to the results of the item measure order calculations as shown in Table 4.

Table 4. Item measure order

Item	Measure	Model S.E.
9	2.40	0.47
3	0.95	0.31
16	0.95	0.31
6	0.72	0.29
13	0.41	0.27
2	0.31	0.24
17	0.25	0.24
14	0.19	0.23
1	0.02	0.22
10	-0.14	0.24
11	-0.14	0.24
7	-0.15	0.22
15	-0.24	0.22
12	-0.34	0.22
5	-0.84	0.19
8	-1.90	0.18
4	-2.47	0.19
Mean	**0.00**	**0.25**
P.SD	**1.06**	**0.06**

Based on Table 4, there are three outlier items (low measurement level) because their item measure values are outside the criteria. The how much criterion is determined by the range from +2SD measure to -2SD measure (Maryati et al., 2019;

Sumintono and Widhiarso, 2015; Wei et al., 2020). Based on these criteria, items 4, 8, and 9 are identified as problematic. Item number 4 is the easiest to answer, while item number 9 is the most difficult to answer. Regarding the precision of item measurement based on model S.E, an item is considered very precise in measurement if the S.E. value is less than 0.5, and not precise if the SE value is greater than 1 (Adams et al., 2022; Sumintono and Widhiarso, 2015). Notably, all items have S.E. values less than 0.5, which means all items have a high level of measurement precision.

How good refers to the content validity of the item, explaining the empirical understanding of the respondents towards the item. How good is determined from the results of the item misfit order analysis as shown in Table 5.

Table 5. Item statistics: misfit order

Item	Outfit MNSQ	Outfit ZSTD	PTMEASURE-AL CORR.
13	3.36	4.60	0.28
12	1.89	2.82	0.51
15	1.23	0.92	0.49
4	1.09	0.46	0.64
6	0.93	-0.09	0.51
7	1.10	0.45	0.45
5	0.95	-0.13	0.23
3	0.83	-0.36	0.48
16	0.83	-0.36	0.48
10	0.83	-0.54	0.46
11	0.83	-0.54	0.46
1	0.69	-1.21	0.43
17	0.64	-1.36	0.59
14	0.60	-1.56	0.49
2	0.54	-1.82	0.53
8	0.38	-3.37	0.64
9	0.27	-2.06	0.07

To assess item validation, several calculation results are considered, such as Outfit MNSQ, Outfit ZSTD value, and Point measure correlation value. These are based on the criteria mentioned in the research method section. Referring to Table 5, the items not meeting the outfit MNSQ criteria are numbers 8, 12, and 13. For the outfit ZSTD value, the items not meeting the criteria are numbers 2, 8, 12, and 13. For the point Measure Correlation Value, the items not meeting the criteria are numbers 5 and 13. Ideally, an item should meet the established criteria for outfit MNSQ, outfit ZSTD, and point measure correlation values. However, if an item does not meet one of these three criteria, it can still be considered and used as an instrument, provided it meets two out of the three established variable value criteria. For example, item number 2 can be used as an instrument because it falls within the acceptable values for outfit MNSQ and point measure correlation. Similarly, item number 5 can be considered as an instrument because it meets the criteria for outfit MNSQ and outfit ZSTD.

Differential item functioning (DIF) in Rasch model analysis refers to the differential item functioning across different groups within the tested population. Simply put, DIF occurs when an item in a test or questionnaire has different levels of difficulty for different subgroups (e.g., based on gender, race, or ethnic group). In this analysis, an illustration of DIF is provided, as shown in Figure 1.

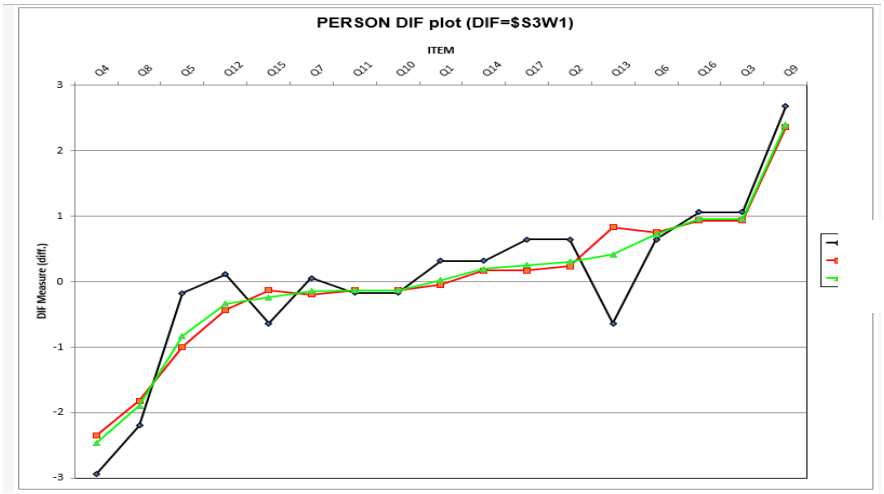

Figure 1. Differential item functioning

Figure 1. shows the differences in DIF measures for each item along the horizontal axis. Three lines represent different groups being compared: the black line with diamonds represents males, the red line with squares represents females, and the green line with triangles represents a combination of males and females. Based on Figure 1, Items 4, 8, 13, and 15 show significantly negative DIF values for the male group compared to the female and combined groups. This means that these items are more difficult for the male group compared to the female group. Conversely, Items 1, 2, 3, 7, 9, 12, 14, and 17 show higher positive DIF values for the male group compared to the female and combined groups. This indicates that these items are easier for the male group. Items with small DIF differences between groups indicate good consistency and no bias towards any group (Anselmi et al., 2019; Rouquette et al., 2019). Meanwhile, items with large DIF differences between groups indicate potential bias, as seen in item 13. Thus, 13 needs to be further examined, modified to ensure fairness in testing or even excluded as an instrument.

A series of analyses of the critical thinking instrument trial results on the topic of wave-particle duality have been conducted using the Rasch model. The results presented by the Rasch model are very comprehensive, as seen from the complex characteristics of each test result. These findings are consistent with previous research using the Rasch model to evaluate measurement instruments in the field of education. (Bond and Fox (2015) demonstrated that the Rasch model is effective in assessing the validity and reliability of instruments, which is also supported by research by Juandi et al., (2023) in the context of critical and reflective thinking skills. In this study, the Rasch model enabled the identification of misfitting items and provided useful information for instrument revision. Additionally, research by Laliyo et al., (2022) emphasized the importance of individual item analysis to understand how each item contributes to the overall construct of the instrument. The findings of this study indicate that the Rasch model can identify biased items while improving the internal consistency of the instrument, ensuring that the instrument can be used more effectively to measure critical thinking skills.

The importance of these findings lies in their contribution to the development of more valid and reliable evaluation instruments in physics education. By using the Rasch model, researchers can ensure that the instruments used truly measure the critical thinking skills required to understand complex physics concepts. The scientific implication of this research is to provide empirical evidence that the Rasch model is an effective tool for evaluating and improving measurement instruments in the context of higher education.

Practically, the results of this research can be used by educators and curriculum developers to improve the quality of evaluation and learning, ensuring that students not only understand physics concepts cognitively but also develop critical thinking skills. These findings also emphasize the importance of continuous revision and the development of instruments. By conducting in-depth analyses of item performance and respondent abilities, researchers can identify areas needing improvement and ensure that the instruments remain relevant and effective in measuring critical thinking skills. In the long term, the use of the Rasch model in educational evaluation can contribute to the overall improvement of education quality, providing a better learning experience for students and preparing them to face real-world challenges.

CONCLUSION

This study analyses the validity and reliability of a critical thinking skills instrument on the topic of wave-particle duality using the Rasch model. Based on the analysis results, it was found that the instrument has excellent item reliability (0.93) and moderate person reliability (0.68), with a Cronbach's alpha value of 0.86 indicating very good internal reliability. The validity of the instrument was tested with infit and outfit MNSQ and ZSTD values, all of which are within the expected acceptance range, indicating that the data fit the Rasch model. Additionally, the construct validity analysis shows that the Rasch model can explain most of the variance in the data, although there are some additional factors not explained by the model. Thus, the critical thinking skills instrument developed on the topic of wave-particle dualism can be used. Of course, this should be done by taking into account several inputs based on the results of the validity and reliability analysis that have been described.

ACKNOWLEDGEMENTS

Thanks very much to Lembaga Pengelola Dana Pendidikan (LPDP), the Indonesian Ministry of Finance, and Hamzanwadi University for providing me with financial support during my study at Universitas Pendidikan Indonesia, Bandung.

REFERENCES

Adams, D., Chuah, K. M., Sumintono, B., and Mohamed, A. (2022). Students' readiness for e-learning during the COVID-19 pandemic in a South-East Asian university: a Rasch analysis. *Asian Education and Development Studies*, *11*(2), 324–339. https://doi.org/10.1108/AEDS-05-2020-0100

Alves, E. G., and Santos, A. L. M. (2021). Photoelectric effect: development of a quantitative experiment. *Revista Brasileira de Ensino de Fisica*, *43*, 1–9. https://doi.org/10.1590/1806-9126-RBEF-2021-0146

Andrich, D., and Marais, I. (2019). A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences. In *Springer*.

https://doi.org/10.1007/978-981-13-7496-8
Anselmi, P., Colledani, D., and Robusto, E. (2019). A Comparison of Classical and Modern Measures of Internal Consistency. *Frontiers in Psychology*, *10*(December), 1–12. https://doi.org/10.3389/fpsyg.2019.02714
Arabatzis, T. (2017). How Physica Became Physics. In *Science & Education*. Science & Education. https://doi.org/https://doi.org/10.1007/s11191-017-9946-7
Bao, L., and Koenig, K. (2019). Physics education research for 21st century learning. *Disciplinary and Interdisciplinary Science Education Research*, *1*(1), 1–12. https://doi.org/10.1186/s43031-019-0007-8
Bond, T. ., and Fox, C. (2015). *Applying the Rasch model fundamental measurement in the human sciences (3rd ed.)*. Routledge.
Boone, W. J., and Staver, J. R. (2020). Correction to: Advances in Rasch Analyses in the Human Sciences. In *Springer*. https://doi.org/10.1007/978-3-030-43420-5_21
Chiang, C. T., Huth, M., Trützschler, A., Kiel, M., Schumann, F. O., Kirschner, J., and Widdra, W. (2015). Boosting laboratory photoelectron spectroscopy by megahertz high-order harmonics. *New Journal of Physics*, *17*. https://doi.org/10.1088/1367-2630/17/1/013035
Connor, M. O., Deeks, H. M., Dawn, E., Metatla, O., Roudaut, A., Sutton, M., Thomas, L. M., Glowacki, B. R., Sage, R., Tew, P., Wonnacott, M., Bates, P., Mulholland, A. J., and Glowacki, D. R. (2018). Sampling molecular conformations and dynamics in a multiuser virtual reality framework. *SCIENCE ADVANCES*, *4*, 1–10.
Daineko, Y., Dmitriyev, V., and Ipalakova, M. (2016). *Using Virtual Laboratories in Teaching Natural Sciences : An Example of Physics Courses in University*. 1–9. https://doi.org/10.1002/cae.21777
Engelhard, G. J. (2013). *Invariant Measurement: Using Rasch Models in the Social Behavioral, and Health Sciences*. Routledge.
Gubsky, D. S., Kleschenkov, S. B., and Mamay, I. V. (2019). Virtual laboratory for microwave measurements. *Comput Appl Eng Educ*, *August*, 1–10. https://doi.org/10.1002/cae.22164
Hadzhikolev, E., Hadzhikoleva, S., Yotov, K., and Orozova, D. (2020). Models for Multicomponent Fuzzy Evaluation, with a Focus on the Assessment of Higher-Order Thinking Skills. *TEM Journal*, *9*(4), 1656–1662. https://doi.org/10.18421/TEM94-43
Handayani, Y., Rahmawati, R., and ... (2023). Using Rasch Model to Analyze Reliability and Validity of Concept Mastery Test on Electricity and Magnetism Topic. *JIPF (Jurnal Ilmu …*, *8*(2), 226–239. https://doi.org/http://dx.doi.org/10.26737/jipf.v8i2.3877
Hermann, G. (2020). The Significance for Natural Philosophy of the Move from Classical to Modern Physics. *Journal for General Philosophy of Science*, *51*(4), 627–629. https://doi.org/10.1007/s10838-020-09533-3
Juandi, T., Kaniawati, I., Samsudin, A., and Riza, L. S. (2023). Implementing the rasch model to assess the level of students' critical and reflective thinking skills on the photoelectric effect. *Momentum: Physics Education Journal*, *7*(2). https://doi.org/https://doi.org/10.21067/mpej.v7i2.8252
Juandi, T., Kaniawati, I., Samsudin, A., Verawati, N. N. S. P., and Zahara, L. (2024). Learning Problems of Prospective Teachers in Lectures Modern Physics During the Covid-19 Pandemic. *KnE Social Sciences*, *2024*, 1221–1238. https://doi.org/10.18502/kss.v9i13.16064
Kanim, S. (2020). Demographics of physics education research. *Physical Review*

Physics Education Research, *16*(2), 20106. https://doi.org/10.1103/PhysRevPhysEducRes.16.020106

Laliyo, L. A. R., La Kilo, A., Paputungan, M., Kunusa, W. R., Dama, L., and Panigoro, C. (2022). Rasch Modelling To Evaluate Reasoning Difficulties, Changes of Responses, and Item Misconception Pattern of Hydrolysis. *Journal of Baltic Science Education*, *21*(5), 817–835. https://doi.org/10.33225/jbse/22.21.817

Lin, C., Griffiths, M. D., and Pakpour, A. H. (2018). Psychometric evaluation of Persian Nomophobia Questionnaire : Differential item functioning and measurement invariance across gender. *Journal of Behavioral Addictions*, *7*(1), 100–108. https://doi.org/10.1556/2006.7.2018.11

Lin, C., Imani, V., Broström, A., Årestedt, K., Pakpour, A. H., and Griffiths, M. D. (2019). Evaluating the Psychometric Properties of the 7-Item Persian Game Addiction Scale for Iranian Adolescents. *Frontiers in Psychology*, *10*(February), 1–13. https://doi.org/10.3389/fpsyg.2019.00149

Ling Lee, W., Chinna, K., and Sumintono, B. (2021). Psychometrics assessment of HeartQoL questionnaire: A Rasch analysis. *European Journal of Preventive Cardiology*, *28*(12), E1–E5. https://doi.org/10.1177/2047487320902322

Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., and Fitrianawati, M. (2021). HOTs Multiple Choice and Essay Questions: A Validated Instrument to Measure Higher-order Thinking Skills of Prospective Teachers. *Journal of Turkish Science Education*, *18*(4), 674–690. https://doi.org/10.36681/tused.2021.97

Maryati, M., Prasetyo, K. un Z., Wilujeng, I., and Sumintoni, B. (2019). Measuring Teachers' Pedagogical Content Knowledge Using Many-Facet Rasch Model. *Cakrawala Pendidikan*, *38*(3), 452–464. https://doi.org/10.21831/cp.v38i3.26598

Mulhayatiah, D., Sinaga, P., Rusdiana, D., Kaniawati, I., and Junissetiawati, D. (2022). Modern Physics E-book Based Multirepresentation for Hybrid Learning. *European Online Journal of Natural and Social Sciences*, *11*(4), 1166–1177. http://www.european-science.com

Nitko, A. J., and Brookhart, S. M. (2014). Educational Assessment of Students: Sixth Edition. In *Pearson New International Edition*. United States of America: Pearson Education.

Nitriani, N., Darsikin, D., and Saehana, S. (2022). Kolb's learning style analysis in solving HOTS questions for prospective physics teacher students. *Momentum: Physics Education Journal*, *6*(1), 59–72. https://doi.org/10.21067/mpej.v6i1.5593

Planinic, M., Boone, W. J., Susac, A., and Ivanjek, L. (2019). Rasch analysis in physics education research : Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 20111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Rahmayanti, H., Ichsan, I. Z., Arif, W. P., Sa'diyah, R., Irwandani, and Fachrial, N. F. H. (2022). Higher-Order Thinking Skills of High School and College Students on Flood Mitigation. *Journal of People, Plants, and Environment*, *25*(1), 33–38. https://doi.org/10.11628/ksppe.2022.25.1.33

Rouquette, A., Hardouin, J., Vanhaesebrouck, A., Se´bille, V., and Coste, J. l. (2019). Differential Item Functioning (DIF) in composite health measurement scale : Recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. *Plos One*, 1–16. https://doi.org/10.1371/journal.pone.0215073

Roza, M., and Festiyed, F. (2023). Kualitas Tes Bermuatan Etnosains Untuk Mengukur Keterampilan Berfikir Kritis Mahasiswa. *Quantum: Jurnal Inovasi Pendidikan Sains*, *14*(2), 206. https://doi.org/10.20527/quantum.v14i2.15599

Samsudin, A., Suhandi, A., Linuwih, S., Kaniawati, I., Fauzi, D. M., Aminudin, A. H., Fratiwi, N. J., and Adimayuda, R. (2023). Preliminary development of simulation on refutational-text (sort) to change conception of physics: Rasch and nvivo analysis. *AIP Conference Proceedings*, *2614*(June). https://doi.org/10.1063/5.0125826

Sinaga, P., and Feranie, S. (2017). *Enhancing Critical Thinking Skills and Writing Skills through the Variation in Non-Traditional Writing Task*. *10*(2), 69–84.

Sumintono, B., and Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Bandung: Trim Komunikata.

Sun, Y., and Latora, V. (2020). The evolution of knowledge within and across fields in modern physics. *Scientific Reports*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-68774-w

Susbiyanto, S., Hidayat, T., Surtikanti, H. K., Riandi, R., Juandi, T., Rochman, S., and Chatib, M. (2024). Perception of Citizen Science Project in Ecology Courses Using Rasch Measurement Model. *KnE Social Sciences*, *2024*, 1159–1169. https://doi.org/10.18502/kss.v9i13.16057

Tarchi, C., and Mason, L. (2019). Effects of critical thinking on multiple-document comprehension. *European Journal of Psychology of Education*, *v35*(n2), p289-313. https://doi.org/https://doi.org/10.1007/s10212-019- 00426-8)

Trilling, B., and Fadel, C. (2009). *21st Century Skills: Learning for Life in Our Times* (Firts). Jossey-Bass.

Verawati, N. N. S. P., Ernita, N., and Prayogi, S. (2022). Enhancing the Reasoning Performance of STEM Students in Modern Physics Courses Using Virtual Simulation in the LMS Platform. *International Journal of Emerging Technologies in Learning*, *17*(13), 267–277. https://doi.org/10.3991/ijet.v17i13.31459

Wei, S., Chee, C., Looi, K., and Sumintono, B. (2020). Assessing computational thinking abilities among Singapore secondary students : a Rasch model measurement analysis. *Journal of Computers in Education*, *0123456789*. https://doi.org/10.1007/s40692-020-00177-2

Widhiarso, W., and Sumintono, B. (2016). Examining response aberrance as a cause of outliers in statistical analysis. *PAID*, *98*, 11–15. https://doi.org/10.1016/j.paid.2016.03.099

Yu, H., Sun, L., Ji, S., Deng, R., Mo, Z., and Xie, Q. (2022). Enhanced Photoelectric Properties of Multilayer Graphene Mg 2 Si/Si Heterojunction Photodetector. *IEEE Photonics Journal*, *14*(1). https://doi.org/10.1109/JPHOT.2021.3128881